

VU Research Portal

Recent advances in extracting and processing rich semantics from medical texts

Denecke, Kerstin; van Harmelen, Frank

published in

Artificial Intelligence in Medicine
2019

DOI (link to publisher)

[10.1016/j.artmed.2018.07.004](https://doi.org/10.1016/j.artmed.2018.07.004)

document version

Publisher's PDF, also known as Version of record

document license

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Denecke, K., & van Harmelen, F. (2019). Recent advances in extracting and processing rich semantics from medical texts. *Artificial Intelligence in Medicine*, 93, 11-12. <https://doi.org/10.1016/j.artmed.2018.07.004>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

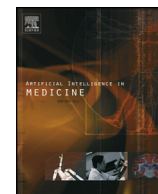
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl



Recent advances in extracting and processing rich semantics from medical texts



1. Introduction

Pharmaceutical companies, healthcare organisations and individual patients exploiting advances in translational medicine and informational infrastructure, are now increasingly recording detailed patient records, an activity that was traditionally limited to only clinical interest. The traditional clinical records already comprised a broad range of clinical documents including nurse letters, discharge summaries and radiology reports describing a patient's health status, diagnoses, applied procedures and observations of the health care team. The rich semantics such as facts, experiences, opinions or information that are hidden in those medical documents can now be combined with further information extracted from background documents such as clinical guidelines, documentation from medical trials and research literature, as well as self-reporting by patients. The semantics from these combined sources could – when extracted automatically – support a broad range of applications including clinical decision support systems, outcome analysis, cohort analysis, etc. Physicians could learn about the experiences of their colleagues, get hints to critical events in the treatment of a specific patient or receive information for improving treatments. Furthermore, such extracted semantics can be used to improve the general healthcare process by updating medical guidelines, identifying unexpected treatment interactions, monitoring healthcare quality, etc. Finally, there is an increased awareness that rich semantics such as sentiments, opinions and other qualitative factors are relevant in ensuring individualized care. Research on this topic has been presented in two workshops held in conjunction with the ESWC 2016 and AIME 2017 conferences. In this special issue we present extended versions of selected workshop papers, but also new research is included.

2. Rich semantics

With rich semantics, we refer to concepts and their relations and characteristics described in written text. Existing methods for information extraction from clinical texts addressed the extraction of occurrences of diagnoses, clinical treatments or medications mainly for the purpose of clinical coding, detection of drug interactions or contraindications. In recent years, approaches gained in interest that go beyond this and that support the extraction of rich semantics. Rich semantics can include descriptions of clinical events, relations among clinical events (e.g. causality relations), but also subjectivity, polarity, emotion or even comparison, for example

- A change in the health status (e.g., a patient can suddenly feel better or worse),
- Critical events, unexpected situations or specific medical conditions that impact the patient's life (e.g., “tumor is malignant” as such is a

fact, but this medical condition is negative for the patient since it might lead to health problems or death),

- The outcome or effectiveness of a treatment (e.g., a surgery can be successfully completed),
- Experiences or opinions towards a treatment or a type of drug (e.g., a patient or a physician can describe serious adverse events after drug consumption),
- The certainty of a diagnosis (e.g., a physician can be uncertain about some diagnosis).

3. Summary of the papers

This special issue focuses on approaches for extracting and processing rich semantics from multiple texts. We solicited submissions that focus on new methods, best practice approaches, lesson learned or evaluation of extraction and processing systems development or identify causes of failures.

Sonntag and Profitlich propose a new integrated decision support system based on textual information extraction, faceted search, and information visualisation [1]. More specifically, their system allows faceted search on structured data and provides a visualisation tool to explore temporal relations between laboratory values and diagnoses. Their system architecture exploits open source tools (UIMA, SOLR) comprising exchangeable modules for search and information extraction. Two use case studies in nephrology and mammography are presented to illustrate the usefulness of the faceted search and visualisation as an integrated decision support application.

Relation classification is a natural language processing task aiming at identifying the relation between two entities in a sentence. This might for example be a relation between a medical problem and a treatment that is applied. Extracting and classifying relations is crucial for many NLP applications such as question answering and knowledge base completion. He, Guan and Dai propose a convolutional neural network (CNN) architecture with a multi-pooling operation for medical relation classification on clinical records and explore a loss function with a category-level constraint matrix [2]. Experiments using the 2010 i2b2/VA relation corpus are reported. They demonstrate that these models outperform previous single-model methods.

Medical social media provide opinions on treatments, on healthcare providers, but also valuable information on experiences with drug consumption and effects of a treatment. Jiménez-Zafra et al. analyzed sentiments in Spanish drug forums to find out how people express their opinion in medical forums [3]. They analyzed the language to determine the best way to tackle sentiment analysis in this domain. They applied supervised learning and lexicon-based sentiment analysis techniques to learn more about features for sentiment analysis and to study the classification accuracy.

In clinical practice, classifying diagnosis and procedures according to medical classification systems is crucial. Traditionally, classification codes are assigned manually or by systems that rely upon concept-based or rule-based classification methods. Such methods can reach their limit easily due to the restricted coverage of handcrafted rules and of the vocabulary in underlying terminological systems. To overcome this limitation, Deng et al. is introducing a processing pipeline for mapping free text to classification codes realized through convolutional neural networks [4]. In particular, they show that enriching the original input by related concepts of a semantic network facilitates semantic matching.

4. Future of extracting rich semantics

To conclude, papers in this special issue cover relevant topics of extracting rich semantics from medical texts. Application of deep learning for extracting or classifying medical data is another topic with increasing popularity. An important question is how to make applications that exploit such rich semantics more reliable and adaptable. Even though data transmission standards such as Health Level 7 and the Clinical Data Interchange Standard CDISC are available in the health-care domain, free text documents are prevalent in daily clinical practice. Thus, research on integrating extracted rich semantics with data transmission standards would be of interest in the future. Beyond this, we have in the medical domain a huge variety of terminological systems, ontologies etc. which could in principle be used for supporting extraction technologies. Research shows that these sources have the potential to improve the quality of the algorithms. Another important topic is the representation and storage of the extracted rich semantics for further analysis, for example with the help of semantic web technologies. There is an increased awareness that rich semantics such as

sentiments, opinions and other qualitative factors are relevant in ensuring individualized care. Rich semantics can in the future be used in clinical decision support systems, with particular support from semantic web technologies. Future work should address these issues in more depth.

Acknowledgements

We would like to thank Carlo Combi for providing the chance to organize this special issue. We also thank the reviewers for their efforts to guarantee the high quality of this special issue. Finally, we thank all the authors who have contributed to this special issue.

References

- [1] Sonntag D, Profitlich H-J. An architecture of open-source tools to combine textual information extraction, faceted search and information visualisation. *Artif Intell Med* 2018.
- [2] He B, Guan Y, Dai R. Classifying medical relations in clinical text via convolutional neural networks. *Artif Intell Med* 2018.
- [3] Jiménez-Zafra Salud María, Martín Valdivia María Teresa, Molina-González M Dolores, Ureña-López L Alfonso. How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain. *Artif Intell Med* 2018.
- [4] Deng Y, Faulstich L, Denecke K. Towards automatic encoding of medical procedures using convolutional neural networks and autoencoders. *Artif Intell Med* 2018.

Kerstin Denecke*

Bern University of Applied Science, Bern, Switzerland

Frank van Harmelen

Vrije Universiteit Amsterdam, Amsterdam, The Netherlands E-mail address: kerstin.denecke@bfh.ch (K. Denecke)

* Corresponding author.